REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
23-11-2015	Final Report	1-Oct-2011 - 31-Jul-2015

4. TITLE AND SUBTITLE
Final Report: Scalable Biomarker Discovery for Diverse High-Dimensional Phenotypes

5a. CONTRACT NUMBER
W911NF-11-1-0429

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Curtis Huttenhower

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Harvard School of Public Health
Biostatistics
President and Fellows of Harvard College
Boston, MA 02115 -6028

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
60119-MA.16

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Historically, most biological and medical investigations have examined a few discrete outcomes of interest, and only a few controllable parameters were modified to perturb or improve these outcomes. Investigations of this form went hand-in-hand with the development of inferential statistics, which provide the quantitative tools to detect which perturbations successfully improve outcomes. Biological and clinical research has entered a realm of modifying hundreds or thousands of experimental parameters in high throughput, however, and high-dimensional statistics have been developed to understand which of these modifications in turn significantly improve outcome

15. SUBJECT TERMS
high-dimensional data, association testing, multiple input multiple output, statistical methods

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Curtis Huttenhower
UU	UU	UU	UU		19b. TELEPHONE NUMBER 617-432-4912

Report Title

Final Report: Scalable Biomarker Discovery for Diverse High-Dimensional Phenotypes

ABSTRACT

Historically, most biological and medical investigations have examined a few discrete outcomes of interest, and only a few controllable parameters were modified to perturb or improve these outcomes. Investigations of this form went hand-in-hand with the development of inferential statistics, which provide the quantitative tools to detect which perturbations successfully improve outcomes. Biological and clinical research has entered a realm of modifying hundreds or thousands of experimental parameters in high throughput, however, and high-dimensional statistics have been developed to understand which of these modifications in turn significantly improve outcome. The field has now reached a point where hundreds or thousands of outcomes can be simultaneously measured as well, but few statistical tools exist to answer the question, "When many experimental or patient outcomes are measured simultaneously, and many experimental parameters or treatments are modified, which modifications are significantly associated with improved outcomes?" This project thus aims to develop novel statistical methods for efficiently associating many controllable predictor variables with many observed response variables with high sensitivity and specificity.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received		Paper
07/30/2012	4.00	Karen E. Nelson, Mihai Pop, Heather H. Creasy, Bruce W. Birren, Richard A. Gibbs, Sarah K. Highlander, George M. Weinstock, Richard K. Wilson, Owen White, Makedonka Mitreva, Donna M. Muzny, John C. Martin, Erica J. Sodergren, James Versalovic, Yanjiao Zhou, Yiming Zhu, Laurie Zoloth, Jeremy D. Zucker, Dana A. Busam, Joseph L. Campbell, Shane R. Canon, Brandi L. Cantarel, Patrick S. Chain, I-Min A. Chen, Lei Chen, Shaila Chhibba, Ken Chu, Dawn M. Ciulla, Jose C. Clemente, Sandra W. Clifton, Sean Conlan, Jonathan Crabtree, Mary A. Cutting, Noam J. Davidovics, Catherine C. Davis, Todd Z. DeSantis, Carolyn Deal, Kimberley D. Delehaunty, Floyd E. Dewhirst, Elena Deych, Yan Ding, David J. Dooling, Shannon P. Dugan, W. Michael Dunne, A. Scott Durkin, Robert C. Edgar, Rachel L. Erlich, Candace N. Farmer, Ruth M. Farrell, Karoline Faust, Michael Feldgarden, Victor M. Felix, Sheila Fisher, Anthony A. Fodor, Larry Forney, Leslie Foster, Valentina Di Francesco, Jonathan Friedman, Dennis C. Friedrich, Catrina C. Fronick, Lucinda L. Fulton, Hongyu Gao, Nathalia Garcia, Georgia Giannoukos, Christina Giblin, Maria Y. Giovanni, Jonathan M. Goldberg, Johannes Goll, Antonio Gonzalez, Allison Griggs, Sharvari Gujja, Brian J. Haas, Holli A. Hamilton, Emily L. Harris, Theresa A. Hepburn, Brandi Herter, Diane E. Hoffmann, Michael E. Holder, Clinton Howarth, Katherine H. Huang, Susan M. Huse, Jacques Izard, Janet K. Jansson, Huaiyang Jiang, Catherine Jordan, Vandita Joshi, James A. Katancik, Wendy A. Keitel, Scott T. Kelley, Cristyn Kells, Susan Kinder-Haake, Nicholas B. King, Rob Knight, Dan Knights, Heidi H. Kong, Omry Koren, Sergey Koren, Karthik C. Kota, Christie L. Kovar, Nikos C. Kyrpides, Patricio S. La Rosa, Sandra L. Lee, Katherine P. Lemon, Niall Lennon, Cecil M. Lewis, Lora Lewis, Ruth E. Ley, Kelvin Li, Asif T. Chinwalla, Ashlee M. Earl, Michael G. FitzGerald, Robert S. Fulton, Kymberlie Hallsworth-Pepin, Elizabeth A. Lobos, Ramana Madupu, Vincent Magrini, Anup A. Mahurkar, Peter J. Mannon, Elaine R. Mardis, Victor M. Markowitz, Konstantinos Mavrommatis, Jamison M. McCorrison, Daniel McDonald, Jean McEwen, Amy L. McGuire, Pamela McInnes, Teena Mehta, Kathie A. Mihindukulasuriya, Jason R. Miller, Patrick J. Minx, Irene Newsham, Chad Nusbaum, Michelle O'Laughlin, Joshua Orvis, Ioanna Pagani, Krishna Palaniappan, Shital M. Patel, Matthew Pearson, Jane Peterson, Mircea Podar, Craig Pohl, Katherine S. Pollard, Margaret E. Priest, Lita M. Proctor, Xiang Qin, Jeroen Raes, Jacques Ravel, Jeffrey G. Reid, Mina Rho, Rosamond Rhodes, Kevin P. Riehle, Maria C. Rivera, Beltran Rodriguez-Mueller, Yu-Hui Rogers, Matthew C. Ross, Carsten Russ, Ravi K. Sanka, Pamela Sankar, J. Fah Sathirapongsasuti, Jeffery A. Schloss, Patrick D. Schloss, Thomas M. Schmidt, Matthew Scholz, Lynn Schriml, Alyxandria M. Schubert, Nicola Segata, Julia A. Segre, William D. Shannon, Richard R. Sharp, Thomas J. Sharpton, Narmada Shenoy, Nihar U. Sheth, Gina A. Simone, Indresh Singh, Chris S. Smillie, Jack D. Sobel, Daniel D. Sommer, Paul Spicer, Granger G. Sutton, Sean M. Sykes, Diana G. Tabbaa, Mathangi Thiagarajan, Chad M. Tomlinson, Manolito Torralba, Todd J. Treangen, Rebecca M. Truty, Tatiana A. Vishnivetskaya, Jason Walker, Lu Wang, Zhengyuan Wang, Doyle V. Ward, Wesley Warren, Mark A. Watson, Christopher Wellington, Kris A. Wetterstrand, James R. White, Katarzyna Wilczek-Boney, Yuan Qing Wu, Kristine M. Wylie, Todd Wylie, Chandri Yandava, Liang Ye, Yuzhen Ye, Shibu Yooseph, Bonnie P. Youmans, Lan Zhang, Aye M. Wollam, Kim C. Worley, Jennifer R. Wortman, Sarah K. Young, Qiandong Zeng, Kjersti M. Aagaard, Olukemi O. Abolude, Emma Allen-Vercoe, Eric J. Alm, Lucia Alvarado, Gary L. Andersen, Scott Anderson, Elizabeth Appelbaum, Harindra M. Arachchi, Gary Armitage, Cesar A. Arze, Tulin Ayvaz, Carl C. Baker, Lisa Begg, Tsegahiwot Belachew, Veena Bhonagiri, Monika Bihan, Martin J. Blaser, Toby Bloom, Vivien R. Bonazzi, Paul Brooks, Gregory A. Buck, Christian J. Buhay, Michelle G. Giglio, Curtis Huttenhower, Dirk Gevers, Joseph F. Petrosino, Sahar Abubucker, Jonathan H. Badger, Barbara A. Methé, Konstantinos Liolios, Bo Liu, Yue Liu, Chien-Chi Lo, Catherine A. Lozupone, R. Dwayne Lunsford, Tessa Madden. A framework for human microbiome research, Nature, (06 2012): 215. doi: 10.1038/nature11209

07/30/2012 3.00 Krishna Palaniappan, Shital M. Patel, Matthew Pearson, Jane Peterson, Mircea Podar, Craig Pohl, Katherine S. Pollard, Doyle V. Ward, Wesley Warren, Mark A. Watson, Christopher Wellington, Kris A. Wetterstrand, James R. White, Katarzyna Wilczek-Boney, YuanQing Wu, Kristine M. Wylie, Todd Wylie, Chandri Yandava, Liang Ye, Yuzhen Ye, Shibu Yooseph, Bonnie P. Youmans, Lan Zhang, Yanjiao Zhou, Yiming Zhu, Laurie Zoloth, Jeremy D. Zucker, Ioanna Pagani, Catrina C. Fronick, Lucinda L. Fulton, Hongyu Gao, Nathalia Garcia, Georgia Giannoukos, Christina Giblin, Maria Y. Giovanni, Jonathan M. Goldberg, Johannes Goll, Antonio Gonzalez, Allison Griggs, Sharvari Gujja, Susan Kinder Haake, Brian J. Haas, Holli A. Hamilton, Emily L. Harris, Katherine P. Lemon, Niall Lennon, Sergey Koren, Cecil M. Lewis, Lora Lewis, Ruth E. Ley, Karthik C. Kota, Christie L. Kovar, Nikos C. Kyrpides, Patricio S. La Rosa, Sandra L. Lee, Mihai Pop, Margaret E. Priest, Lita M. Proctor, Xiang Qin, Jeroen Raes, Jacques Ravel, Jeffrey G. Reid, Mina Rho, Rosamond Rhodes, Kevin P. Riehle, Maria C. Rivera, Ramana Madupu, Vincent Magrini, John C. Martin, Makedonka Mitreva, Donna M. Muzny, Erica J. Sodergren, James Versalovic, Aye M. Wollam, Kim C. Worley, Jennifer R. Wortman, Sarah K. Young, Qiandong Zeng, Kjersti M. Aagaard, Olukemi O. Abolude, Emma Allen-Vercoe, Eric J. Alm, Lucia Alvarado, Gary L. Andersen, Scott Anderson, Elizabeth Appelbaum, Harindra M. Arachchi, Gary Armitage, Cesar A. Arze, Tulin Ayvaz, Carl C. Baker, Lisa Begg, Tsegahiwot Belachew, Veena Bhonagiri, Monika Bihan, Martin J. Blaser, Toby Bloom, Vivien Bonazzi, J. Paul Brooks, Gregory A. Buck, Christian J. Buhay, Dana A. Busam, Joseph L. Campbell, Shane R. Canon, Brandi L. Cantarel, Patrick S. G. Chain, I-Min A. Chen, Lei Chen, Shaila Chhibba, Ken Chu, Dawn M. Ciulla, Jose C. Clemente, Sandra W. Clifton, Sean Conlan, Jonathan Crabtree, Mary A. Cutting, Noam J. Davidovics, Catherine C. Davis, Todd Z. DeSantis, Carolyn Deal, Kimberley D. Delehaunty, Floyd E. Dewhirst, Elena Deych, Yan Ding, David J. Dooling, Shannon P. Dugan, Wm Michael Dunne, A. Scott Durkin, Robert C. Edgar, Rachel L. Erlich, Candace N. Farmer, Ruth M. Farrell, Karoline Faust, Michael Feldgarden, Victor M. Felix, Sheila Fisher, Anthony A. Fodor, Larry J. Forney, Leslie Foster, Valentina Di Francesco, Jonathan Friedman, Dennis C. Friedrich, Beltran Rodriguez-Mueller, Yu-Hui Rogers, Matthew C. Ross, Carsten Russ, Ravi K. Sanka, Pamela Sankar, J. Fah Sathirapongsasuti, Jeffery A. Schloss, Patrick D. Schloss, Thomas M. Schmidt, Matthew Scholz, Lynn Schriml, Alyxandria M. Schubert, Nicola Segata, Julia A. Segre, William D. Shannon, Richard R. Sharp, Thomas J. Sharpton, Narmada Shenoy, Nihar U. Sheth, Gina A. Simone, Indresh Singh, Christopher S. Smillie, Jack D. Sobel, Daniel D. Sommer, Paul Spicer, Granger G. Sutton, Sean M. Sykes, Diana G. Tabbaa, Mathangi Thiagarajan, Chad M. Tomlinson, Manolito Torralba, Todd J. Treangen, Rebecca M. Truty, Tatiana A. Vishnivetskaya, Jason Walker, Lu Wang, Zhengyuan Wang, Konstantinos Liolios, Bo Liu, Kelvin Li, Yue Liu, Chien-Chi Lo, Catherine A. Lozupone, R. Dwayne Lunsford, Tessa Madden, Anup A. Mahurkar, Peter J. Mannon, Elaine R. Mardis, Victor M. Markowitz, Konstantinos Mavromatis, Jamison M. McCorrison, Daniel McDonald, Jean McEwen, Amy L. McGuire, Pamela McInnes, Teena Mehta, Kathie A. Mihindukulasuriya, Jason R. Miller, Patrick J. Minx, Irene Newsham, Chad Nusbaum, Michelle O'Laughlin, Joshua Orvis, Theresa A. Hepburn, Brandi Herter, Diane E. Hoffmann, Michael E. Holder, Clinton Howarth, Katherine H. Huang, Susan M. Huse, Jacques Izard, Janet K. Jansson, Huaiyang Jiang, Catherine Jordan, Vandita Joshi, James A. Katancik, Wendy A. Keitel, Scott T. Kelley, Cristyn Kells, Nicholas B. King, Dan Knights, Heidi H. Kong, Omry Koren, Bruce W. Birren, Richard A. Gibbs, Sarah K. Highlander, Barbara A. Methé, Karen E. Nelson, Joseph F. Petrosino, George M. Weinstock, Richard K. Wilson, Owen White, Curtis Huttenhower, Dirk Gevers, Rob Knight, Sahar Abubucker, Jonathan H. Badger, Asif T. Chinwalla, Heather H. Creasy, Ashlee M. Earl, Michael G. FitzGerald, Robert S. Fulton, Michelle G. Giglio, Kymberlie Hallsworth-Pepin, Elizabeth A. Lobos.
Structure, function and diversity of the healthy human microbiome,
Nature, (06 2012): 207. doi: 10.1038/nature11234

08/30/2013 9.00 Timothy L Tickle, Harry Sokol, Dirk Gevers, Kathryn L Devaney, Doyle V Ward, Xochitl C Morgan, Joshua A Reyes, Samir A Shah, Neal LeLeiko, Scott B Snapper, Athos Bousvaros, Joshua Korzenik, Bruce E Sands, Ramnik J Xavier, Curtis Huttenhower. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment,
Genome Biology, (04 2012): 0. doi: 10.1186/gb-2012-13-9-r79

10/09/2014 10.00 Michelle G Rooks, Patrick Veiga, Leslie H Wardwell-Scott, Timothy Tickle, Nicola Segata, Monia Michaud, Carey Ann Gallini, Chloé Beal, Johan ET van Hylckama-Vlieg, Sonia A Ballal, Xochitl C Morgan, Jonathan N Glickman, Dirk Gevers, Curtis Huttenhower, Wendy S Garrett. Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission,
The ISME Journal, (02 2014): 0. doi: 10.1038/ismej.2014.3

10/09/2014	11.00	Dirk Gevers, Subra Kugathasan, Lee A. Denson, Yoshiki Vázquez-Baeza, Will Van Treuren, Boyu Ren, Emma Schwager, Dan Knights, Se Jin Song, Moran Yassour, Xochitl C. Morgan, Aleksandar D. Kostic, Chengwei Luo, Antonio González, Daniel McDonald, Yael Haberman, Thomas Walters, Susan Baker, Joel Rosh, Michael Stephens, Melvin Heyman, James Markowitz, Robert Baldassano, Anne Griffiths, Francisco Sylvester, David Mack, Sandra Kim, Wallace Crandall, Jeffrey Hyams, Curtis Huttenhower, Rob Knight, Ramnik J. Xavier. The Treatment-Naive Microbiome in New-Onset Crohn's Disease, Cell Host & Microbe, (03 2014): 0. doi: 10.1016/j.chom.2014.02.005

TOTAL: **5**

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received	Paper	
11/13/2015	12.00	Xochitl C Morgan, Boyko Kabakchiev, Levi Waldron, Andrea D Tyler, Timothy L Tickle, Raquel Milgrom, Joanne M Stempak, Dirk Gevers, Ramnik J Xavier, Mark S Silverberg, Curtis Huttenhower. Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease, , (04 2015): 67. doi:
11/13/2015	13.00	Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen AM, Peet A, Tillmann V, Pöhö P8, Mattila , Lähdesmäki H, Franzosa EA, Vaarala O11, de Goffau M, Harmsen H, Ilonen J, Virtanen SM, Clish CB, Oreši? M, Huttenhower C, Knip M, DIABIMMUNE Study Group, Xavier RJ. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes., , (02 2015): 260. doi:
11/13/2015	14.00	Dan Knights, Mark S Silverberg, Rinse K Weersma, Dirk Gevers, Gerard Dijkstra, Hailiang Huang, Andrea D Tyler, Suzanne van Sommeren, Floris Imhann, Joanne M Stempak, Hu Huang, Pajau Vangay, Gabriel A Al-Ghalith, Caitlin Russell, Jenny Sauk, Jo Knight, Mark J Daly, Curtis Huttenhower, Ramnik J Xavier. Complex host genetics influence the microbiome in inflammatory bowel disease, [object Object], (12 2014): 107. doi:
11/14/2015	15.00	Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, Baldassano RN, Noe JD, Rosh J, Markowitz J, Heyman MB, Griffiths AM, Crandall WV, Mack DR, Baker SS, Huttenhower C, Keljo DJ, Hyams JS, Kugathasan S, Walters TD, Aronow B, Xavier RJ, Gevers D, Denson LA. Pediatric Crohn's disease patients exhibit specific ileal transcriptome and microbiome signature. , Clinical Investgation, (08 2014): 3617. doi:

TOTAL: **4**

Number of Papers published in non peer-reviewed journals:

(c) Presentations

1- "High-precision functional profiling of the gut microbiome for characterization during inflammatory disease." University of Pittsburgh Immunology seminar. Pittsburgh, PA, 2015
2- "Microbial communities in the Boston MBTA mass transit system." MetaSUB First International Summit on Metagenomics and Metadesign of Subways and Urban Biomes. New York, NY, 2015
3- "Towards systems-level functional profiling of microbial communities and the human microbiome." University of Pennsylvania Microbiology seminar. Philadelphia, PA, 2015
4- "High-precision Functional Profiling of Microbial Communities and the Human Microbiome." 41st Annual Northeast Bioengineering Conference. Albany, NY, 2015
5- "High-precision functional profiling of microbial communities and the human microbiome." Simons Foundation Symposium on Genomics in Single Cells and Microbiomes. New York, NY, 2015
6- "A Tour of the bioBakery: Computational Tools for Microbial Community Analysis." Broad Institute Medical and Population Genetics seminar. Cambridge, MA, 2015 (presented by Eric Franzosa)
7- "The human microbiome and biomarker discovery." Harvard CATALYST Understanding Biomarker Science workshop. Boston, MA, 2015
8- "Towards systems-level functional profiling of microbial communities and the human microbiome." Channing Division of Network Medicine Theodore L. Badger Lecture. Boston, MA, 2015
9- "Characterizing the gut microbial ecosystem for diagnosis and therapy in inflammatory bowel disease." Keystone Symposium on Gut Microbiota Modulation of Host Physiology. Keystone, CO, 2015
10- "The microbiome in IBD and analysis methods for microbial communities." International Inflammatory Bowel Disease Genetics Consortium meeting. Barcelona, Spain, 2015
11- "An Introduction to Microbial Community Analyses." Evomics and Genomics workshop. Cesky Krumlov, Czech Republic, 2015
12- "High-specificity methods for profiling microbial communities and the human microbiome." University of Oregon Computer Science colloquium. Eugene, OR, 2014
13- "Metagenomics, metatranscriptomics, and multi'omic integration." Massachusetts General Hospital Center for the Study of Inflammatory Bowel Disease research symposium. Boston, MA, 2014
14- "High-precision methods for metagenomic and metatranscriptomic profiling." New York University Medical School seminar. New York, NY, 2014
15- "Gut microbial epidemiology and biogeography." University of Washington Genome Sciences seminar. Seattle, WA, 2014
16- "High-precision profiling of microbial communities and the human microbiome." University of Oregon Institute for Theoretical Sciences seminar. Eugene, OR, 2014
17- "Computational Approaches for the Human Microbiome in Health and Disease," 12th Biennial Congress of the Anaerobe Society of the Americas. Chicago, IL, 2014
18- "An introduction to the microbiome and quantitative methods for microbial community analysis," HSPH Biostatistics Summer Program in Quantitative Sciences. Boston, MA, 2014
19- "An introduction to the microbiome and methods for microbial community analysis," Harvard/MIT Minority Introduction to Engineering, and Science. Boston, MA, 2014
20- "An introduction to metagenomics," Strategies and Techniques for Analyzing Microbial Population Structure. Woods Hole, MA, 2014
21- "Host-microbiome transcriptional crosstalk and clinical outcome in a large ileal pouch-anal anastomosis (IPAA) cohort," 109th International Titisee Conference. Titisee, Germany, 2014
22- "Microbiome Bioinformatics Tools: A Tutorial," Keystone Symposium on Exploiting and Understanding Chemical Biotransformations in the Human Microbiome. Big Sky, MO, 2014 (presented by Xochitl Morgan)
23- "A Tour of the BioBakery: Computational Tools for Microbial Community Analysis," Harvard School of Public Health Program in Quantitative Genomics Short Course series. Boston, MA, 2014 (presented by Eric Franzosa)
24- "High-precision functional profiling and integration of metagenomes and metatranscriptomes," Weizmann Institute Systems Biology Seminar Series. Rehovot, Israel, 2013
25- "Bug bytes: bioinformatics for the human microbiome in health and disease," University of Michigan Molecular and Clinical Epidemiology of Infectious Diseases (MAC-EPID) symposium. Ann Arbor, MI, 2013
26- "Computational methods for meta'omic characterization of the human microbiome," Tufts Computer Science Department Colloquium Series. Medford, MA, 2013
27- "Interaction of Host Gene Expression and the Human Gut Microbiome in Pouchitis," INFORMS Annual Meeting. Minneapolis, MN, 2013 (presented by Levi Waldron)
28- "Adding depth to human microbiome studies with multi'omic data integration," International Human Microbiome Congress. Hangzhou, China, 2013
29- "Bug bytes: bioinformatics for the human microbiome in health and disease," Canadian Student Health Research Forum. Alberta, Canada, 2013
30- "From microbes to microbiota and back: using thousands of genomes to understand thousands of metagenomes," Harvard School of Public Health Bioinformatics Core Forum. Boston, MA, 2013
31- "From microbes to microbiota and back: using thousands of genomes to understand thousands of metagenomes," Symposium and Workshop on New Methods for Phylogenomics. Austin, TX, 2013
32- "Computational methods for meta'omic characterization of the human microbiome," Los Alamos National Laboratory Center for Nonlinear Studies seminar. Los Alamos, NM, 2013
33- "An introduction to metagenomics," Strategies and Techniques for Analyzing Microbial Population Structure. Woods Hole, MA, 2013

34- "Bug bytes: Computational analysis methods for microbial communities," University of Oregon BioBE center seminar. Eugene, OR, 2013
35- "From microbial surveys to mechanisms of interaction in the human microbiome," University of Colorado at Boulder BioFrontiers Institute seminar. Boulder, CO, 2013
36- "Detailing the human microbiome with meta'omics," New England Primate Research Center. Southboro, MA, 2012
37- "Computational methods for meta'omic characterization of the human microbiome," Forsyth Institute seminar. Cambridge, MA, 2012
38- "Computational methods for meta'omic characterization of the human microbiome," Procter and Gamble BioFusion Symposium. Cincinnati, OH, 2012
39- "Bug bytes: Computational analysis methods for microbial communities," Army Research Office workshop on the skin microbiome. Boulder, CO, 2012
40- "Meta'omic Characterization of Microbial Community Function in Health and Disease," Mount Sinai School of Medicine Department of Health Evidence and Policy Grand Rounds. New York, NY, 2012
41- "Bug bytes: Computational analysis methods for microbial communities," Carnegie Mellon Lane Center for Computational Biology seminar. Pittsburgh, PA, 2012
42- "Meta'omic characterization of microbial community function in health and disease," American Society for Microbiology Conference on Beneficial Microbes. San Antonio, TX, 2012
43- "Bug bytes: bioinformatics for metagenomics and microbial community analysis," Lewis-Sigler Institute for Integrative Genomics seminar. Princeton, NJ, 2012
44- "Bug bytes: bioinformatics for metagenomics and microbial community analysis," 8th International Purdue Symposium on Statistics. West Lafayette, IN, 2012
45- "Bug bytes: computational methods for microbial community analysis," Woods Hole Marine Biology Laboratory seminar. Woods Hole, MA, 2012
46- "Computational tools for functional analysis of microbial communities," Cloud Computing for the Microbiome Workshop. Boulder, CO, 2012

Number of Presentations: 46.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received	Paper

TOTAL:

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received		Paper
08/30/2013	6.00	Dan Knights, Mark S. Silverberg, Rinse K. Weersma, Dirk Gevers, Gerard Dijkstra, Hailiang Huang, Andrea D. Tyler, Suzanne van Sommeren, Floris Imhann, Joanne M. Stempak, Caitlin Russell, Jenny Sauk, Jo Knight, Mark J. Daly, Curtis Huttenhower, Ramnik J. Xavier. Complex host genetics drives bacterial dysbiosis in inflammatory bowel disease, IN submission (04 2013)
08/30/2013	7.00	Sanne P. Smeekens, Curtis Huttenhower, Anca Riza, Frank L. van de Veerdonk, Patrick L.J.M. Zeeuwen, Joost Schalkwijk, Jos W.M. van der Meer, Ramnik J. Xavier, Mihai G. Netea, Dirk Gevers. Skin Microbiome Imbalance in Patients with STAT1/STAT3 Defects Impairs Innate Host Defense Responses, Journal of Innate immunity (06 2013)
08/30/2013	8.00	Michelle G. Rooks, Patrick Veiga, Leslie H. Wardwell-Scott, Nicola Segata, Timothy Tickle, Carey Ann Gallini, Chloe Beal, Monia Michaud, Johan E. T. van Hylckama-Vlieg, Jonathan N. Glickman, Dirk Gevers, Curtis Huttenhower, Wendy S. Garrett. Composition and function of the gut microbiome in experimental colitis during active disease and treatment-induced remission, IN submission (04 2013)
TOTAL:		**3**

Number of Manuscripts:

Books

Received		Book
TOTAL:		

Received	Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

1- ISCB Overton Prize (Harvard School of Public Health, 2015)
2- eLife Sponsored Presentation Series early career award (Harvard School of Public Health, 2014)
3- Presidential Early Career Award for Scientists and Engineers (Harvard School of Public Health, 2012)

Graduate Students

NAME	PERCENT_SUPPORTED	Discipline
Boyu Ren	0.95	
Emma Schwager	0.10	
FTE Equivalent:	**1.05**	
Total Number:	**2**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
Timothy Tickle	0.20
FTE Equivalent:	**0.20**
Total Number:	**1**

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Curtis Huttenhower	0.08	No
FTE Equivalent:	**0.08**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	1.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:	1.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	1.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME	
Boyu Ren	
Total Number:	**1**

Names of personnel receiving PHDs

NAME
Total Number:

Names of other research staff

NAME	PERCENT_SUPPORTED
Yo-Sup Moon	0.85
George Weingart	0.30
FTE Equivalent:	**1.15**
Total Number:	**2**

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

See Attachment

Technology Transfer

HAllA's implementation as a Python package (see http://huttenhower.sph.harvard.edu/halla) has been carried out in collaboration with Weingart Informatics, an independent software development contractor in San Francisco. This has allowed academic development and validation of the algorithm to be carried out efficiently by students and postdoctoral fellows, while Dr. Weingart has provided industry-quality code, unit testing, packaging, documentation, and distribution. He has begun expanding this implementation to a generalizable Python platform for scientific workflow execution, AnADAMA, which may be a target for future industry partnership in the lab.

60119-MA: Scalable biomarker discovery for diverse high-dimensional phenotypes

Associate Professor Curtis Huttenhower, Department of Biostatistics, Harvard School of Public Health

Our final technical report for this project includes completion of all methodological development for HAllA (the Hierarchical All-against All input/output association testing approach), in addition to work previously completed for the MaAsLin (Multivariate Analysis with Linear models) system for high-dimensional biomarker discovery in compositional data. Both are available as open-source software packages with documentation, demonstration data, and tutorials at http://huttenhower.sph.harvard.edu/halla and http://huttenhower.sph.harvard.edu/maaslin, respectively. Our final publication list includes six manuscripts (PMIDs 22699609, 22699610, 24629344, 23949665, 23013615, and 25732063) and two currently in review (HAllA and its application to the oral microbiome).

Problem Statement

Historically, most biological and medical investigations have examined a few discrete outcomes of interest, and only a few controllable parameters were modified to perturb or improve these outcomes. Investigations of this form went hand-in-hand with the development of inferential statistics, which provide the quantitative tools to detect which perturbations successfully improve outcomes. Biological and clinical research has entered a realm of modifying hundreds or thousands of experimental parameters in high throughput, however, and high-dimensional statistics have been developed to understand which of these modifications in turn significantly improve outcome. The field has now reached a point where hundreds or thousands of outcomes can be simultaneously measured as well, but few statistical tools exist to answer the question, "When many experimental or patient outcomes are measured simultaneously, and many experimental parameters or treatments are modified, which modifications are significantly associated with improved outcomes?" This project thus aims to develop novel statistical methods for efficiently associating many controllable *predictor* variables with many observed *response* variables with high sensitivity and specificity (**Fig. 1**).

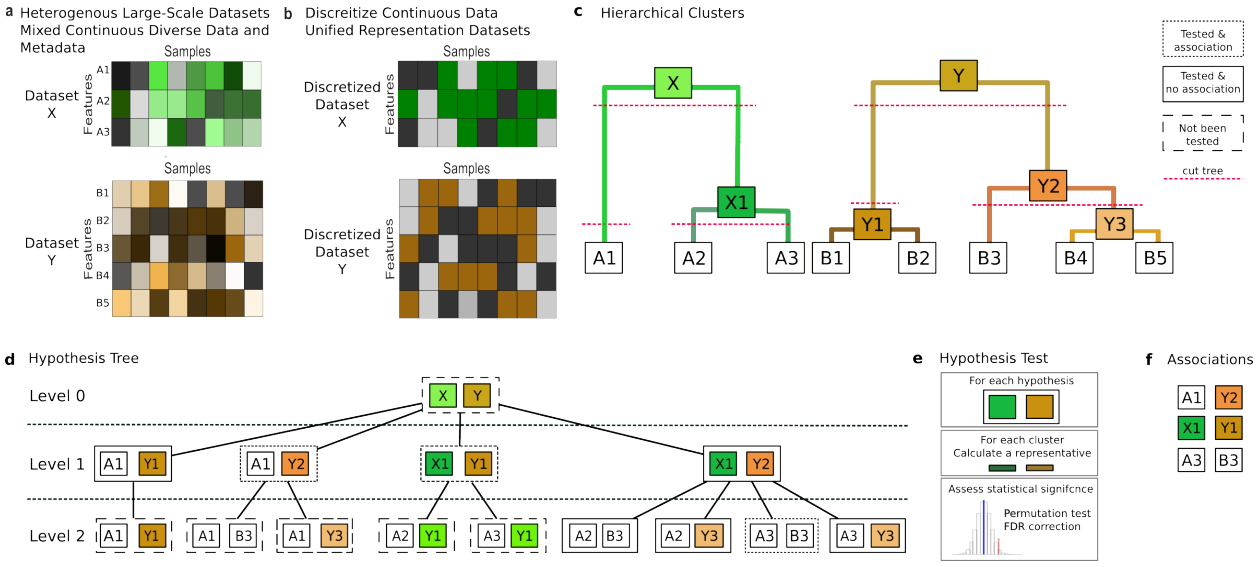

Figure 1: Overview of HAllA. A) Two or more input datasets are represented in matrix form as features (rows) and samples (columns). B) Continuous data are (optionally) discretized to provide a unified representation of potentially heterogeneous feature types. C) Features within each data set are single linkage hierarchically clustered, using normalized mutual information as the default, fully generalizable similarity metric. D) A hypothesis tree is built by coupling clusters between two datasets at equivalent relative depths. Each hypothesis node has compares two

clusters, with all pairs of children of the two clusters forming the next level of hypothesis testing. E) Hypothesis testing is performed by, first, selecting each cluster's medoid as a representative summary (optionally either multiple correspondence analysis or principle component analysis instead), and a permutation test is then used to determine which pairs are significantly associated between the two datasets. F) Significant associations are reported after false discovery rate controlling for each hypotheses set (family, level, or all).

Results Summary

Generalized multiple input/output association testing: the Hierarchical All-against-All approach

HAllA (Hierarchical All-against-All association testing) is a novel statistical method for well-powered association discovery in high-dimensional heterogeneous datasets, which we have developed, implemented, validated, and applied to diverse datasets. It combines hierarchical hypothesis testing with false discovery rate correction over highly generalizable association measures, yielding high-powered discovery of linear and non-linear patterns in categorical or continuous high-dimensional data. Data and metadata to be associated are hierarchically clustered for dimensionality reduction, and nonparametric permutation testing identifies relationships between the resulting blocks of correlated features.

HAllA was validated and optimized with synthetic data, outperforming exhaustive all-against-all association testing and alternative similarity measures such as the Maximum Information Coefficient and Spearman correlation in Types I and II error and in runtime (**Fig. 2**). The recommended HAllA algorithm first groups features by single-linkage agglomerative hierarchical clustering using normalized mutual information (NMI), then compares blocks of features by descending each tree logarithmically. Outliers are removed by comparing the relative dissimilarities within and between datasets, after which cluster medoids are compared (also using NMI) and significant associations determined by permutation testing. False discovery rates are globally controlled while maintaining power by correcting within each tree level. Each choice in the algorithm is modularized and configurable, however, allowing users to select other measures more appropriate for homogeneous data (e.g. continuously valued Spearman correlations) or to use other block summarization techniques (e.g. MCA or PCA). The software implementation and documentation are available at http://huttenhower.sph.harvard.edu/halla.

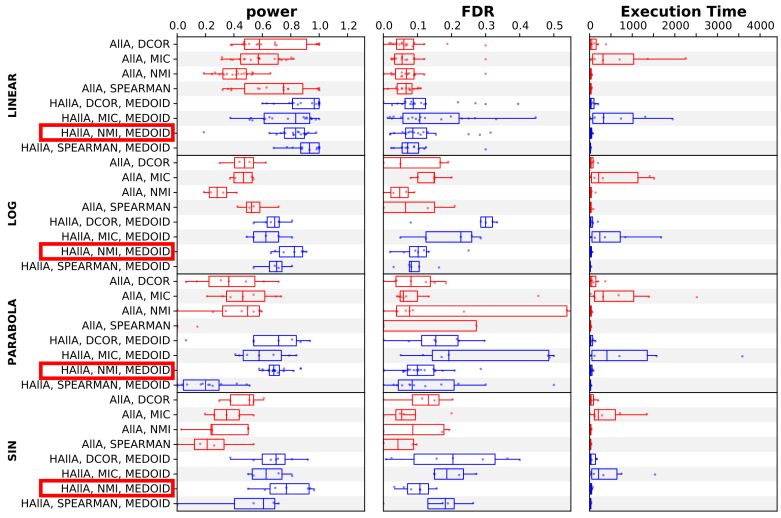

Figure 2: HAllA efficiently maintains Type I and II error relative to alternative approaches. We simulated 30 datasets with features associated by linear, parabolic, and sinusoidal relationships. HAllA outperformed exhaustive all-against-all (AllA) testing using multiple measures: the maximum information coefficient (MIC), normalized mutual information (NMI) alone, and alternative summarizations including MCA or PCA.

We created the Python package STRUDEL (Stratified Rudimentary Data Exploration) to produce synthetic data with defined correlation structure within datasets and associations between blocks of variables among datasets. Each evaluation dataset contained correlated

features, and each paired dataset contained correlations between blocks of features with both linear and non-linear patterns. Using these simulated data, we evaluated the performance of HAllA and naive all-against-all (AllA) methods using NMI, MIC, distance correlation (dcor), and Spearman correlation as similarity metrics. Simple correlations identify only linear associations, while dcor requires continuous data and performed poorly for discovery of complex nonlinear associations. MIC was very time-consuming even for small datasets often had a high FDR. However, NMI efficiently found all types of associations (linear, parabolic, and sinusoidal) and when applied using hierarchical testing improved both precision and recall. As a negative control, applying HAllA to null datasets containing no associations resulted in the expected baseline false discovery rate (a user configurable parameter).

We applied HAllA to three real-world datasets containing highly diverse data types from several application areas. First, we assessed data published by Martin et al (Hepatology 2007) for 21 liver lipid levels and 120 genes' hepatic transcription levels in 40 wild-type and peroxisome proliferator-activated receptor-α (PPARα)-deficient mice. HAllA recovered all associations previously reported by Gonzalez et al (J. Stat. Software 2008) using canonical correlation analysis, including for example transcriptional activation of xenobiotic metabolism genes Cyp3a11 and CAR1 in conjunction with high fatty acid levels. We further identified novel associations between a cluster of transcripts including CAR1 and ACOTH (fatty acid transport and trafficking) and a cluster of fatty acids including C18.0, C20.3n.6, and C20.4n.6. The associations identified by HAllA were a strict superset of those previously reported, demonstrating the utility of the method as a general-purpose tool that can find general patterns with high sensitivity (**Fig. 3**).

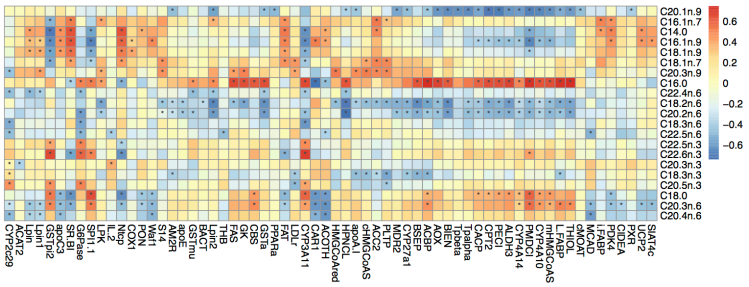

Figure 3: HAllA identifies validated and novel associations between mouse liver transcripts and fatty acid levels. Associations between hepatic fatty acids and gene expression in data from Martin et al (Hepatology 2007). '*' indicates significant associations identified by HAllA between corresponding genes and fatty acids, a strict superset of those previously reported.

Next, HAllA expanded upon microbe-metabolite associations reported the infant gut microbiome by Kostic et al (CHM 2015). This test of the hygiene hypothesis studied a prospective cohort of 960 Finnish, Russian, and Estonian infants at risk of type 1 diabetes. These subjects were followed for three years, while monthly stool samples and a variety of clinical metadata (e.g. breastfeeding, diet, allergies) were collected. The resulting 104 stool samples were assessed for microbial community composition by 16S rRNA gene sequencing and profiled metabolomically. We applied HAllA to the abundances of 20 genera and 284 metabolites from these data, again identifying a set of relationships that were previously detected (e.g. *Veilonella* and sphyingomyelins 22:0, 24:0, 24:1, C16:0, and C18:0; *Ruminococcus* and sphingomyelins 24:0 and 24:1). We further recovered an association between *Haemophilus* and phenylalanine in conjunction with a novel grouping with tryptophan and tyrosine. *Blautia spp.* were non-linearly associated with long-chain triglycerides, whereas *Ruminococcus spp.* associated with short-chain triglycerides and *Lactobacillus spp.* with both. Finally, *Enterococcus* was positively associated with diacyl and triacyl glycerols, in agreement with independent reports of *Enterococcus faecium* bioactivity in improving diabetic lipid (Roselino Lipids Health 2012) and triglyceride (Cavallini Lipids Health 2009) levels.

Finally, we applied HAllA to data collected from 204 ileal resection patients in which microbial community profiles and host transcriptomes were assessed from 255 biopsies (Morgan CHM 2015). Host transcription was assayed by Affymetrix microarray and microbiome profiles by 16S rRNA gene sequencing. Previous work correlated antibiotic use, inflammation, biopsy location, and clinical outcome with the transcriptome and microbiome, but extensive dimensionality reduction was required to preserve power while determining microbe-transcript associations. We applied HAllA to a highly-abundant (above 50th percentile) and variant subset (above 95th percentile), comprising 108 microbes and 498 transcripts. We found 576 associations between features and feature clusters; 34 included at least one of the top gene principal component loadings described previously. Of these, 21 corresponded to the genes most influential in the previously reported principal component 9, which represented only 1% of expression variation but was correlated with the most microbes.

Additionally, we found that the abundance of *Escherichia* was correlated with host expression of MUC1 and nitric oxide synthase. Expression of the sodium / bile acid transporter SLC10A2, the primary bile transporter in the distal ileum, was correlated with the genera *Coprococcus*, *Blautia*, and *Clostridium* from the family Erysipelotrichaceae. Notably, antibiotic use was associated with increased SLC10A2 expression and total bile acids, in agreement with Miyata et al (J. Pharma. Exp. Therapy 2011), and both Crohn's disease and ulcerative colitis have previously been associated with decreased metagenomic abundance of bile salts among the Lachnospiraceae and Erysipelotrichaceae (Labbe PLoS ONE 2014).

The manuscript describing HAllA is currently in review, and it has previously been presented at the 2015 Broad Institute retreat, the 2015 Statistical and Applied Mathematical Sciences Institute workshop on Discovering Patterns in Human Microbiome Data, the 2014 Intelligent Systems for Molecular Biology (ISMB) conference, and the 2014 Dana-Farber Cancer Insistute Biostatistics and Computational Biology seminar series. Its software and evaluation package are complete, documented, and supported by online tutorials and an active user group (see http://huttenhower.sph.harvard.edu/halla and http://bitbucket.org/biobakery/halla). It is currently in use by ongoing collaborations with Dr. Jessica Green at the University of Oregon, Dr. Frank Nestle (King's College for the MAARS study), and Dr. Mihai Netea (Radboud University Nijmegen). Postdoctoral fellow Dr. Gholamli Rahnavard is the current lead, in collaboration with research associate Dr. Eric Franzosa and completing work previously executed by former research assistant Yo Sup Moon and postdoctoral fellows Timothy Tickle (currently at the Broad Institute) and Levi Waldron (currently at Hunter College).

Multiple input/output microbial community methods: Multivariate Analysis with Linear models

The Multivariate Analysis with Linear models (MaAsLin) method performs efficient high-dimensional association testing specifically for microbial communities. Microbial taxonomic and functional profiles possess unique statistical characteristics, in particular a combination of sparsity (many zero values), compositionality (fixed per-sample sum) or count data, and longitudinal variability (i.e. repeated measures), including typical epidemiological characteristics of heterogeneous high-dimensional metadata. MaAsLin was developed as a well-powered statistical association methodology and software appropriate for microbial data in conjunction with any experimental design and associated phenotypic, clinical, environmental, or 'omic metadata.

The MaAsLin algorithm combines four main steps to associate relative abundance profiles with heterogeneous metadata (**Fig. 4**). First, data are preprocessed to remove outliers, handle missing values, and ensure quality control and consistency. Next, one of three link functions appropriate for microbial profiles is applied: a variance-stabilizing arcsin square-root transformation by default, or a log-linear Gaussian link, or a negative Binomial count link. The first (arcsin-sqrt) is an approximation that is less theoretically appropriate than the latter two, but performs comparably well in evaluations and is substantially more computationally efficient and numerically stable. Third, dimensionality is reduced if necessary using a feature selection step (boosting by default, LASSO and univariate selection optionally available). Finally, significant associations are identified using a mixed effects linear model (zero-inflated by default).

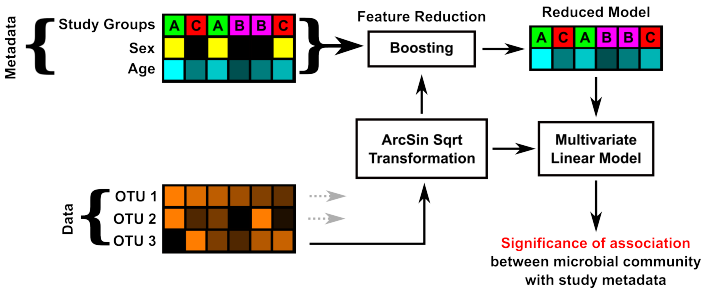

Figure 4: Multivariate linear models (MaAsLin) for high-dimensional relative abundance profile association with multivariate metadata. Multivariate Association by Linear models (MaAsLin) uses a dimensionality-reduced mixed effects model to identify significant covariation between microbes or microbial functions and sample metadata (phenotype, environment, etc.)

MaAsLin has been applied in a variety of published and ongoing studies during the total project period:

- Morgan XC, Kabakchiev B, Waldron L, Tyler AD, Tickle TL, Milgrom R, Stempak JM, Gevers D, Xavier RJ, Silverberg MS, Huttenhower C. "Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease." Genome Biol. 2015 Apr 8;16(1):67
- Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen AM, Peet A, Tillmann V, Pöhö P, Mattila I, Lähdesmäki H, Franzosa EA, Vaarala O, de Goffau M, Harmsen H, Ilonen J, Virtanen SM, Clish CB, Orešič M, Huttenhower C, Knip M; DIABIMMUNE Study Group, Xavier RJ. "The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes." Cell Host Microbe, 2015 Feb 11;17(2):260-73
- Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, Baldassano RN, Noe JD, Rosh J, Markowitz J, Heyman MB, Griffiths AM, Crandall WV, Mack DR, Baker SS, Huttenhower C, Keljo DJ, Hyams JS, Kugathasan S, Walters TD, Aronow B, Xavier RJ, Gevers D, Denson LA. "Pediatric Crohn's disease patients exhibit specific ileal transcriptome and microbiome signature." J Clin Invest, 2015 Mar 2;125(3):1363
- Knights D, Silverberg MS, Weersma RK, Gevers D, Dijkstra G, Huang H, Tyler AD, van Sommeren S, Imhann F, Stempak JM, Huang H, Vangay P, Al-Ghalith GA, Russell C, Sauk J, Knight J, Daly MJ, Huttenhower C, Xavier RJ. "Complex host genetics influence the microbiome in inflammatory bowel disease." Genome Med, 2014 Dec 2;6(12):107
- Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, Schwager E, Knights D, Song SJ, Yassour M, Morgan XC, Kostic AD, Luo C, González A, McDonald D, Haberman Y, Walters T, Baker S, Rosh J, Stephens M, Heyman M, Markowitz J, Baldassano R, Griffiths A, Sylvester F, Mack D, Kim S, Crandall W, Hyams J, Huttenhower C, Knight R, Xavier RJ. "The treatment-naive microbiome in new-onset Crohn's disease." Cell Host and Microbe, 2014 Mar 12;15(3):382-92

- Rooks MG, Veiga P, Wardwell-Scott LH, Tickle T, Segata N, Michaud M, Gallini CA, Beal C, van Hylckama-Vlieg JE, Ballal SA, Morgan XC, Glickman JN, Gevers D, Huttenhower C, Garrett WS. "Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission." ISME J, 2014 Feb. 6
- Smeekens SP, Huttenhower C, Riza A, van de Veerdonk FL, Zeeuwen PL, Schalkwijk J, van der Meer JW, Xavier RJ, Netea MG, Gevers D. " Skin Microbiome Imbalance in Patients with STAT1/STAT3 Defects Impairs Innate Host Defense Responses." J. Innate Imm. 2013
- Huttenhower C, Gevers D, Knight R, The Human Microbiome Project Consortium, White O. "Structure, function and diversity of the healthy human microbiome." Nature, 2012 486(7402):207-14
- Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, The Human Microbiome Project Consortium, White O. "A framework for human microbiome research." Nature, 2012 486(7402):215-21
- Morgan XC*, Tickle TL*, Sokol H*, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. "Dysfunction of the Intestinal Microbiome in Inflammatory Bowel Disease and Treatment." Genome Biology, 2012, 13:R80
- Tommi Vatanen, Aleksandar D. Kostic, Eva d'Hennezel, Heli Siljander, Eric A. Franzosa, Moran Yassour, Raivo Kolde, Hera Vlamakis, Anu-Maaria Hämäläinen, Aleksandr Peet, Vallo Tillmann, Raivo Uibo, Sergei Mokurov, Natalya Dorshakova, Jorma Ilonen, Suvi M. Virtanen, Susanne J. Szabo, Jeff Porter, Harri Lähdesmäki, Curtis Huttenhower, Dirk Gevers, Thomas W. Cullen, Mikael Knip, on behalf of the DIABIMMUNE Study Group, and Ramnik J. Xavier. "From Metagenomics to Mechanism: LPS Immunogenicity Links the Infant Microbiome to Autoimmunity." in revision
- Tiffany Hsu, Regina Joice, Jose Vallarino, Galeb Abu-Ali, Erica M. Hartmann, Afrah Shafquat, Casey DuLong, Catherine Baranowski, Dirk Gevers, Jessica L. Green, Xochitl C. Morgan, John D. Spengler, Curtis Huttenhower. "Urban transit system microbial communities differ by surface type and interaction with humans and environment." in review
- Daniela Börnigen, Boyu Ren, Robert Pickard, Jingfeng Li, Erica M. Hartmann Weihong Xiao, Timothy Tickle, Jennifer Rider, Dirk Gevers, Mary Ellen Davey, Curtis Huttenhower, Maura Gillison. "Alterations in the oral microbiome associated with oral cancer risk factors and their contributions to pathogenesis." in review

MaAsLin has also been presented in a wide range of venues, including:

- "High-precision functional profiling of the gut microbiome for characterization during inflammatory disease." University of Pittsburgh Immunology seminar. Pittsburgh, PA, 2015
- "Microbial communities in the Boston MBTA mass transit system." MetaSUB First International Summit on Metagenomics and Metadesign of Subways and Urban Biomes. New York, NY, 2015
- "Towards systems-level functional profiling of microbial communities and the human microbiome." University of Pennsylvania Microbiology seminar. Philadelphia, PA, 2015
- "High-precision Functional Profiling of Microbial Communities and the Human Microbiome." 41st Annual Northeast Bioengineering Conference. Albany, NY, 2015
- "High-precision functional profiling of microbial communities and the human microbiome." Simons Foundation Symposium on Genomics in Single Cells and Microbiomes. New York, NY, 2015

- "A Tour of the bioBakery: Computational Tools for Microbial Community Analysis." Broad Institute Medical and Population Genetics seminar. Cambridge, MA, 2015 (presented by Eric Franzosa)
- "The human microbiome and biomarker discovery." Harvard CATALYST Understanding Biomarker Science workshop. Boston, MA, 2015
- "Towards systems-level functional profiling of microbial communities and the human microbiome." Channing Division of Network Medicine Theodore L. Badger Lecture. Boston, MA, 2015
- "Characterizing the gut microbial ecosystem for diagnosis and therapy in inflammatory bowel disease." Keystone Symposium on Gut Microbiota Modulation of Host Physiology. Keystone, CO, 2015
- "The microbiome in IBD and analysis methods for microbial communities." International Inflammatory Bowel Disease Genetics Consortium meeting. Barcelona, Spain, 2015
- "An Introduction to Microbial Community Analyses." Evomics and Genomics workshop. Cesky Krumlov, Czech Republic, 2015
- "High-specificity methods for profiling microbial communities and the human microbiome." University of Oregon Computer Science colloquium. Eugene, OR, 2014
- "Metagenomics, metatranscriptomics, and multi'omic integration." Massachusetts General Hospital Center for the Study of Inflammatory Bowel Disease research symposium. Boston, MA, 2014
- "High-precision methods for metagenomic and metatranscriptomic profiling." New York University Medical School seminar. New York, NY, 2014
- "Gut microbial epidemiology and biogeography." University of Washington Genome Sciences seminar. Seattle, WA, 2014
- "High-precision profiling of microbial communities and the human microbiome." University of Oregon Institute for Theoretical Sciences seminar. Eugene, OR, 2014
- "Computational Approaches for the Human Microbiome in Health and Disease," 12th Biennial Congress of the Anaerobe Society of the Americas. Chicago, IL, 2014
- "An introduction to the microbiome and quantitative methods for microbial community analysis," HSPH Biostatistics Summer Program in Quantitative Sciences. Boston, MA, 2014
- "An introduction to the microbiome and methods for microbial community analysis," Harvard/MIT Minority Introduction to Engineering, and Science. Boston, MA, 2014
- "An introduction to metagenomics," Strategies and Techniques for Analyzing Microbial Population Structure. Woods Hole, MA, 2014
- "Host-microbiome transcriptional crosstalk and clinical outcome in a large ileal pouch-anal anastomosis (IPAA) cohort," 109th International Titisee Conference. Titisee, Germany, 2014
- "Microbiome Bioinformatics Tools: A Tutorial," Keystone Symposium on Exploiting and Understanding Chemical Biotransformations in the Human Microbiome. Big Sky, MO, 2014 (presented by Xochitl Morgan)
- "A Tour of the BioBakery: Computational Tools for Microbial Community Analysis," Harvard School of Public Health Program in Quantitative Genomics Short Course series. Boston, MA, 2014 (presented by Eric Franzosa)
- "High-precision functional profiling and integration of metagenomes and metatranscriptomes," Weizmann Institute Systems Biology Seminar Series. Rehovot, Israel, 2013

- "Bug bytes: bioinformatics for the human microbiome in health and disease," University of Michigan Molecular and Clinical Epidemiology of Infectious Diseases (MAC-EPID) symposium. Ann Arbor, MI, 2013
- "Computational methods for meta'omic characterization of the human microbiome," Tufts Computer Science Department Colloquium Series. Medford, MA, 2013
- "Interaction of Host Gene Expression and the Human Gut Microbiome in Pouchitis," INFORMS Annual Meeting. Minneapolis, MN, 2013 (presented by Levi Waldron)
- "Adding depth to human microbiome studies with multi'omic data integration," International Human Microbiome Congress. Hangzhou, China, 2013
- "Bug bytes: bioinformatics for the human microbiome in health and disease," Canadian Student Health Research Forum. Alberta, Canada, 2013
- "From microbes to microbiota and back: using thousands of genomes to understand thousands of metagenomes," Harvard School of Public Health Bioinformatics Core Forum. Boston, MA, 2013
- "From microbes to microbiota and back: using thousands of genomes to understand thousands of metagenomes," Symposium and Workshop on New Methods for Phylogenomics. Austin, TX, 2013
- "Computational methods for meta'omic characterization of the human microbiome," Los Alamos National Laboratory Center for Nonlinear Studies seminar. Los Alamos, NM, 2013
- "An introduction to metagenomics," Strategies and Techniques for Analyzing Microbial Population Structure. Woods Hole, MA, 2013
- "Bug bytes: Computational analysis methods for microbial communities," University of Oregon BioBE center seminar. Eugene, OR, 2013
- "From microbial surveys to mechanisms of interaction in the human microbiome," University of Colorado at Boulder BioFrontiers Institute seminar. Boulder, CO, 2013
- "Detailing the human microbiome with meta'omics," New England Primate Research Center. Southboro, MA, 2012
- "Computational methods for meta'omic characterization of the human microbiome," Forsyth Institute seminar. Cambridge, MA, 2012
- "Computational methods for meta'omic characterization of the human microbiome," Procter and Gamble BioFusion Symposium. Cincinnati, OH, 2012
- "Bug bytes: Computational analysis methods for microbial communities," Army Research Office workshop on the skin microbiome. Boulder, CO, 2012
- "Meta'omic Characterization of Microbial Community Function in Health and Disease," Mount Sinai School of Medicine Department of Health Evidence and Policy Grand Rounds. New York, NY, 2012
- "Bug bytes: Computational analysis methods for microbial communities," Carnegie Mellon Lane Center for Computational Biology seminar. Pittsburgh, PA, 2012
- "Meta'omic characterization of microbial community function in health and disease," American Society for Microbiology Conference on Beneficial Microbes. San Antonio, TX, 2012
- "Bug bytes: bioinformatics for metagenomics and microbial community analysis," Lewis-Sigler Institute for Integrative Genomics seminar. Princeton, NJ, 2012
- "Bug bytes: bioinformatics for metagenomics and microbial community analysis," 8th International Purdue Symposium on Statistics. West Lafayette, IN, 2012

- "Bug bytes: computational methods for microbial community analysis," Woods Hole Marine Biology Laboratory seminar. Woods Hole, MA, 2012
- "Computational tools for functional analysis of microbial communities," Cloud Computing for the Microbiome Workshop. Boulder, CO, 2012

The software is currently available with documentation and demonstration data at http://huttenhower.sph.harvard.edu/maaslin, with an additional Galaxy interface online at http://huttenhower.sph.harvard.edu/galaxy. The project is currently led by postdoctoral fellow Dr. Ayshwarya Subramanian, with previous contributions by former postdoctoral fellows Dr. Soumya Bannerjee (currently at Children's Hospital) and Dr. Timothy Tickle (currently at the Broad Institute), Ph.D. student Emma Schwager, and undergraduate research assistant Yiren Lu.